

DOCUMENT RESUME

ED 223 003

EA 015 164

AUTHOR Gersten, Russell
TITLE Direct Instruction with Special Education Students: A Review of Evaluation Research.
PUB DATE 81
NOTE 33p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Behavior Modification; Classroom Techniques; *Disabilities; Educational Innovation; Educational Research; Elementary Education; *Evaluation Methods; Predictor Variables; Program Effectiveness; *Program Evaluation; Research Problems; *Teaching Methods

IDENTIFIERS *Direct Instruction

ABSTRACT

This paper reviews both the content and the methodological issues involved in the small but growing number of empirical studies on the use of Direct Instruction programs and teaching procedures with special populations. The research reviewed includes experimental and quasi-experimental work, summative evaluations of entire programs, and fine-grained evaluations of single instructional components. These studies evaluate the effects of Direct Instruction reading and language programs among subjects ranging from those with mild academic deficits to the severely handicapped. Studies are organized for discussion according to the type of evaluation design utilized in order to acquaint the reader with the benefits and drawbacks associated with each type of design. For each study reviewed, the major findings are presented and methodological problems and measurement issues analyzed. Directions for future research are suggested. In its final section, the paper turns to the problem of adequately defining independent variables, noting that precision in these definitions is particularly important. Suggestions are drawn from the research for resolving this problem. (Author/PGD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED223003

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Russell Gersten

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Direct Instruction with Special Education Students:
A Review of Evaluation Research

1981

Russell Gersten
University of Oregon

Running Head: Direct Instruction with Special Education Students

ED 015 167

Direct Instruction with Special Education Students

Direct Instruction with Special Education Students

A Review of Evaluation Research

The last five years have witnessed a growing realization that many of the principles and concepts of direct instruction have immediate relevance for special education (Bellamy, Horner, & Inman, 1979; Bateman & Carnine, 1977; Stephens & Rosenshine, 1981; Reith, Polsgrove, & Semmel, 1982; Gersten & Maggs, in press). The term Direct Instruction evolved out of the field of compensatory education, where the Direct Instruction Follow Through model (Becker, Engelmann, Carnine, & Rhine, 1981) for teaching reading, language, and mathematics to low income students in the primary grades has had significant positive effects (cf. Becker, 1977; Stebbins et al., 1977; Carnine & Gersten, 1981). Rosenshine (1976) used the phrase direct instruction in a more generic fashion to describe the set of instructional variables that are consistently related to high levels of academic growth for students in the elementary grades. Rosenshine's definition, unlike the one used in this paper, does not deal with curriculum or issues in instructional design.

Direct Instruction shares many features with the task analytic, behavioral approaches commonly utilized in special education, viz. a belief in the utility of structured curricula materials, a concern with reinforcement of appropriate responses, the modelling and shaping of correct responses, the use of task analysis, and continuous assessment of student performance. Yet there are certain features that distinguish Direct Instruction from most other behavioral approaches. These include: (a) the explicit teaching of "general case" problem solving strategies whenever possible; (b) an emphasis on small group instruction as opposed to students working alone; (c) a systematic technology of correction procedures; (d) principles for cumulative review of previously learned material, and (e) insistence on mastery of each step in the learning process. Specific curriculum programs

in reading, math, and language have been developed that utilize the above features.

This paper reviews the small but growing empirical support of the use of Direct Instruction programs and teaching procedures with special populations. The research conducted thus far has included experimental and quasi-experimental work, summative evaluations of an entire program's effectiveness, and fine-grained evaluations of the importance of one instructional component within an entire program. The subjects range from those with mild academic deficits to the severely handicapped.

Evaluation research is a relatively new venture in education (Cooley & Lohnes, 1976), and particularly in special education where the focus on individualization of educational programs and the wariness of standardized tests often create strong resistance to typical evaluation designs (Kennedy, Note 1). As I confronted the wide range of studies and documentation activities, it seemed as important to highlight conceptual problems and technical difficulties encountered by the investigators in establishing evaluation designs as to share the results.

Consequently, this paper focuses on both the content of the studies and methodological issues in the field of evaluation research. The published studies to be reviewed evaluate the effects of Direct Instruction reading and language programs on a wide range of handicapping conditions. Studies are organized according to type of evaluation design utilized, rather than by population served. This is done so that the reader can understand the unique benefits--and drawbacks--associated with each type of design. For each study reviewed, the major findings are presented, along with an analysis of methodological problems and measurement issues inherent in the design. Where possible, suggestions for future research are offered.

The final section of this essay begins to confront an issue that often hides in the closet in evaluation research--the definition of the independent variable(s). In future research on Direct Instruction, more clarity and precision is needed in defining exactly what is being evaluated. Some suggestions from current research on implementation of structured curriculum programs (Hall & Loucks, 1977; Gersten, Carnine, & Williams, 1982), explanatory observational studies (Leinhardt, Zigmond, & Cooley, 1981), and formative evaluation are presented as promising directions.

Experimental Designs

The strongest evaluation design is one in which students are randomly assigned to one of two educational programs, and the mean performance of the two samples on a battery of academic measures is compared to determine whether one performs at a significantly different level from the other. For this type of study to offer any valuable information, the tests used must be valid, reliable, and sensitive to the goals of the instructional programs (House, Glass, McLean, & Walker, 1978). Also, it is extremely important that the researcher monitor to what extent the new program (in this case, Direct Instruction) is really being implemented in the experimental classes (Gall, 1977; Charters & Jones, 1973). There should also be some monitoring of the comparison classrooms (Becker, Gersten, & Carnine, Note 2) to insure that these classes are not using the experimental teaching method. If all these conditions are met, and the sample size is large enough (usually at least 15 children per condition), one can be reasonably sure that differences between the samples on the post-tests are due to the educational program.

Yet there are so many factors mitigating against this type of design--refusal of parents and school personnel to randomly assign students to

instructional programs, attrition of students, difficulties in finding tests sensitive to the instructional program. Despite the obstacles, three such experimental studies have been conducted with Direct Instruction and handicapped learners; they will be discussed in some detail.

Beginning Reading Instruction with High Risk Students. The first randomized experiment conducted in this field was an investigation of three instructional procedures for "high risk" first graders in a suburban community. Serwer, Shapiro, and Shapiro (1973) deemed 62 kindergarteners as "high risk" on the basis of "observed difference between potential and achievement with due regard in both those areas to the verbal, performance and visual motor dimensions" (p. 242). Screening was based on teacher questionnaires, two reading tests, and two IQ tests. Note that over 20 percent of the suburban students were targeted as "high risk," though typical estimates of students needing some special education services are in the range of 10 to 12 percent.

Students were randomly assigned to one of four remediation groups-- Direct Instruction (30 minutes a day of Distar Reading), Indirect Instruction (use of perceptual motor activities for 30 minutes a day, based on the work of Kephart (1964)), both treatments Combined (15 minutes a day of each treatment), or Control (no special treatment). All students received the remedial treatment in addition to their regular first grade reading program. Another anomaly is that students were tested midyear on a variety of auditory-visual, fine motor, and gross motor tests, and were then given "prescriptions." For the remainder of the year, "The Direct [Instruction] group combined the Distar program ... [with training in]... deficit modalities as diagnosed by using letter forms and sounds and word forms and sounds." The use of modality training is clearly not part of the

Direct Instruction approach (see Bateman, 1979). In this study, the Direct method does not adhere to many principles of Direct Instruction, and the results may be confounded by the unspecified reading programs that the students received an additional hour per day. It is also unclear exactly what features of each method were included in the Combined Treatment. The fidelity of implementation was not measured for any of the treatments, because the special features of each treatment were not specified, and the degree to which the features were actually carried out was not mentioned.

The only significant findings were: Students in the Kephart program performed significantly higher on two non-academic tests of motor skills (Locomotor Balance, Eye-hand Coordination) related to the goals of the Kephart program. They also performed significantly higher on two academically oriented tests (Handwriting, Math) that were not an emphasis of any treatment used in the study. Students in the Direct group performed significantly better on one of the reading tests, the Wrong Endings subtest of the Gates-McKillop.

It is unclear what to make of this study. Despite the myriad of tests administered and statistical tests performed, few significant results were found. In fact, the few significant findings could be due merely to chance. The large standard deviations on many of the measures would seem to indicate that the tests rarely assess the goals of the programs and/or that the population is very heterogeneous.

This flawed study exemplifies the importance of:

- a) offering as clear a definition of the target population as possible;
- b) carefully defining and implementing the educational treatments.
- c) measuring fidelity of implementation of each educational program.
- d) selection of tests that meet the objectives of the programs.

These issues surface to varying degrees in all the studies discussed below.

Language Instruction for the Moderately Retarded. Maggs and Morath (1976) investigated the effects of Distar Language I (a beginning level Direct Instruction language program) on moderately to severely retarded children in a state institution. Twenty-eight students, ages 6 to 14, with Stanford-Binet IQ's between 20 and 45, were randomly assigned to either a Direct Instruction group (one hour per day of Distar Language I coupled with precision teaching procedures) or a comparison group (one hour per day with the Peabody Language Kit supplemented by teacher-generated language programs). The program lasted two full years. At the conclusion of the program, the experimental group scored significantly higher on the Stanford Binet; the mean gain in "mental age" for this group was 22 1/2 months. Since the students were in the program for the 24 months, this approximated normal intellectual growth rates. The control group gained only 7 1/2 mental age months in the 24 months of the program.

Children were also tested on a series of cognitive tasks based on the experimental work of Piaget and Bruner. The authors used these tasks to explore whether concepts learned purely by verbal instruction and the printed page (as in Distar) would transfer to a situation in which the concepts had to be demonstrated by manipulating three-dimensional, concrete objects. The results showed a significantly better post-test performance by the experimental subjects, indicating their skills generalized beyond mere rote learning.

This field study documented that much of the intelligent behavior assessed in tests like the Stanford Binet can be taught through extensive systematic instruction, even to moderate and severely retarded children. A flaw in this study is that no measures of fidelity of implementation/

were included. Children in the Maggs and Morath (1976) study for two years approached the normal level of growth. This point will be returned to when quasi-experimental designs are discussed.

Teaching Reading to Learning Disabled Students. Lloyd et al. (1980, 1981) randomly assigned 23 learning disabled students in the intermediate grades to one of three different classrooms. The two experimental classes utilized remedial Direct Instruction programs in reading (Engelmann et al., 1978) to teach word attack skills and reading comprehension. The teaching techniques described in the teacher's manuals were supplemented with a monitoring system devised by the staff, involving specific probe sheets for academic skill areas. Arithmetic was taught with a traditional basal series.

Students in the comparison classroom received individual and small group instruction in various academic areas (reading, language arts, math, spelling) as well as training in psychological processes believed to be related to learning. The program was considered representative of traditional special education programs for the learning disabled at that time. Students in both experimental and comparison classes were in a full day program.

Students were assessed on: (a) the Wide Range Achievement Test (Jastak & Jastak, 1965), which measures word recognition and arithmetic (computation); (b) the Gilmore Oral Reading Test (Gilmore & Gilmore, 1968); and (c) the Slosson Intelligence Test (Slosson, 1963), which assesses general language competence and comprehension.

Results indicated statistically significant differences favoring the experimental group in measures of both word recognition and comprehension.

The difference in decoding accuracy was significant, $t_{(21)} = 1.68$, $p < .05$. The magnitude of effect was .64 pooled SD units. Significant effects were also found on the Gilmore Accuracy Index and Comprehension Index and the Slosson Intelligence Test. On the other hand, the difference in WRAT arithmetic was clearly non-significant; $t_{(21)} = .658$. Since no Direct Instruction math program was implemented, this offers further confirmation of the effectiveness of the Direct Instruction programs (Campbell & Fiske, 1959).

This study documents the efficacy of Direct Instruction reading programs in decoding and comprehension. It is one of the few experimental studies to use learning disabled students, and the only study to examine the effectiveness of the newly developed comprehension programs. On the technical side, this is one of the best designed special education research studies--with its use of random assignment of children to treatment, the broad range of measures used, its relatively clear definition of the handicapped, and its assessment of performance in both target (reading) and non-target (arithmetic) academic areas. Lloyd et al.'s (1981) use of improvised "standard" scores on the Gilmore is a bit questionable, since there is considerable controversy concerning the use of grade equivalent scores (Horst, Tallmadge, & Wood, 1975) and mental age-to-chronological age ratios (Matarazzo, 1972). Yet, since they used identical scoring procedures for both experimental and comparison groups, there is no reason to assume the procedure systematically biased the results.

Two issues raised by the authors deserve further elaboration and discussion. Though the experimental group consistently surpassed the comparison group, one year of Direct Instruction reading was not sufficient to bring those students to a "normal" level of performance. (That

would represent mean standard scores of 100 on the measures). In fact, there is evidence of a small drop in measured IQ for the two experimental classrooms--from 90.4 to 87.1 for one group and 87.5 to 85.6 for the other. But this drop is appreciably less than the drop for the comparison group. The drop may be artifactual--students were pre-tested on the WISC, which contains performance and language items, and post-tested on the Slosson, which is primarily verbal and hence may underestimate scores of language deficient students. As the authors state, "these data do not argue that... the procedures cured learning disabilities (in ten months), but they do indicate a direction that clearly merits further direction" (p. 31).

A methodological issue that needs to be further pursued with a more finegrained analysis is exactly what the instructional program was like. It is described as "a hybrid behavioral model" using both behavioral principles of programming and direct instruction curricula and teaching techniques. But the study mentions no check on the fidelity of treatment--are the programs implemented as the authors intended, and if not, do the changes that are part of the "hybrid behavioral model" help or hinder student performance? The explanatory observational studies of Leinhardt, Zigmond, and Cooley (1981) offer a methodology for beginning to look at these issues.

Quasi-experimental Designs

The studies discussed above, which involved random assignment of subjects to educational treatment, though imperfect, offer some degree of methodological rigor to the questions they pose. Unfortunately, it is often impossible or infeasible to randomly assign children to educational programs. There are a host of ethical and logistical reasons why most school systems will not allow selection of educational programs for handicapped students to be determined by the toss of a coin or a flip through a

random numbers table. Even Lloyd and Maggs, principal authors of the two previous studies, have been unable to achieve random assignment in subsequent research (Lloyd, Note 3; Maggs, Note 4).

One alternative to an experimental design is a quasi-experimental design. Three major types of quasi-experimental designs appear to be appropriate for evaluation research with special education populations:

- a) Non-equivalent comparison groups (Cook & Campbell, 1979),
- b) Norm-referenced comparisons (Becker & Engelmann, 1976; Tallmadge, 1977),
- c) Time-series designs (Cook & Campbell, 1979).

Each has its limitations; yet, used with care, these can be useful designs; and they have already provided some valuable information on the effects of Direct Instruction programs on special education students.

Non-equivalent Comparison Groups: Issues in Selection of Beginning Reading Curricula. A recent study by L'E. Stein and Goldman (1980) contrasted the effectiveness of two phonics-based programs--Distar and Palo Alto--on 63 "primary grade children with "reading problems" between the ages of 6 and 8. The children's IQs were in the normal range (mean = 100.1). All the children were formally diagnosed as possessing minimal brain dysfunction, but the authors (rightly, in my opinion) felt that "the term was not particularly relevant to each specific child's curriculum, which was based on pretest needs as measured through academic assessment" (p. 53).

L'E. Stein and Goldman did a careful job of articulating the two treatment programs and why they were selected. Both programs are phonics-based instruction for beginning readers since the bulk of research supports phonics approaches to beginning reading instruction (Chall, 1979; Plaum et al., 1980). The major differences between Distar and Palo Alto, according

to the authors, are:

1. Distar insists on complete mastery of each component skill in the program before continuing on, while Palo Alto calls for less complete mastery.

2. In the initial stages of instruction, Distar teaches phonetic decoding strategies only whereas Palo Alto relies on both systematic phonics and contextual cues (Gill, 1974).

3. Palo Alto allows for greater teacher latitude in directing child behavior than the scripted Distar lessons.

4. On a more theoretical level, Palo Alto is heavily rooted in cognitive theories that look at the teacher as a facilitator of active learning (Gibson & Levin, 1975); in Distar, the teacher directs all stages in the learning process (Becker et al., 1981; Stephens & Rosenshine, 1981).

No details were given on how students were assigned to each program. Thus, as with all non-equivalent group designs, there is always a fear that the two samples were not equivalent at the beginning of treatment, and the results somewhat biased. However, the researchers scrupulously attempted to insure that the groups were comparable and no significant differences were found on entry scores on either the reading recognition or reading comprehension subtests of the Peabody Individual Achievement Test (PIAT) or measured IQs.

Students were post-tested on both these subtests of the PIAT and a significant difference was found between Distar and Palo Alto: $t = 5.27$, $df = 56.5$; $p < .001$. The mean gain for the Distar group corresponded to 15 months for 9 months of instruction; the mean gain for the Palo Alto students was approximately 7 months. The differences were present in both reading

recognition and comprehension; though it should be mentioned that the early level of the PIAT measures primarily literal (as opposed to inferential) comprehension skills.

The study looks at two commonly used special education curriculum programs. It articulates four salient variables of interest (mentioned above). And the study gives a reasonably clear indication of the combination of variables that was most effective. It is unfortunate that no classroom observations were done during the teaching sessions to assess: (a) the extent to which teachers were following the procedures specified by the curriculum programs, and (b) quantitative differences in the aforementioned four variables of interest. Another crucial issue would be amount of practice on each skill (including review and delayed practice exercises that each program allots). The importance of practice and review has been documented by Good and Grouws (1979). Procedures for this type of research are discussed in Meyer (in press) and Gersten, Carnine, and Williams (1982).

The L'E. Stein and Goldman (1980) study is the only study involving Direct Instruction to utilize the non-equivalent comparison group design. In this case, the authors did much to insure that the two samples were comparable.

Norm-Referenced Comparisons: Longitudinal Growth of Moderately Retarded Children. Norm-referenced comparisons are relatively new designs originating in the work of Horst, Tallmadge, and Wood (1975) in Title I evaluations. In many ways, it is an easy design to implement. Students are pretested and post-tested on the same form and level of a well-normed, reliable standardized achievement test, whose content matches the objectives of

their curriculum program. The gain against the standardization sample is assessed for statistical and educational significance. There is no need to select a control group; the standardization sample of the test serves this purpose. In theory, the standardization sample should be similar to the target population; i.e., Title I students should be compared to other low income students, special education students to other handicapped students, etc. (Tallmadge, 1977). In reality, virtually every standardized test in use is normed on a primarily non-handicapped population; evaluators have no choice but to use these. The situation is slowly changing with the development of several new instruments. A curriculum-referenced test of academic and social skills called The Student Progress Record (Oregon Mental Health Division, 1977) has been developed in the state of Oregon that has been normed exclusively on a trainable mentally retarded (TMR) population. The Adaptive Performance Inventory (White, 1980) was developed and standardized exclusively with severely handicapped preschoolers. The Trainee Performance Sample (Irvin, Gersten, Bellamy, Taylor, & Close, 1981), a vocational skill test, was normed exclusively on a sample of severely retarded adults.

Although norm-referenced comparisons are not often used in special education, their potential contribution in the future may be great. In most situations, however, one has only the non-handicapped standardization samples to use as comparison standards, and if they are too stringent a standard to use. For example, in the two studies discussed previously by Maggs and Morath (1976) and Lloyd et al. (1980), there are slight, non-significant drops against the norm group.

There is another example of a norm-referenced comparison that I wish to discuss. Gersten and Maggs (in press) used a norm-referenced evaluation

design to assess the cognitive and academic progress over five years of a sample of twelve adolescents and preadolescents whose initial IQs would place them in the high moderate range of retardation. The students were engaged in the Distar Language programs for the entire five years. After 18 months of language instruction, when they had mastered the basic language competencies and ability to follow instructions in an academic setting, they began Distar Reading I. Each lesson (reading or language) lasted 30 minutes per day and was supplemented by 30 minutes a day of independent seatwork. At the end of the five years the students had completed most of the lessons in the Language III program and most had begun the Reading III program. (The Level III programs contain material intended for average third graders.)

Since no comparison group was feasible in this situation, the standardization sample of the Stanford-Binet served as a comparison standard; the progress of these students was compared to the progress of those in the Stanford-Binet norm group with similar IQs. Students were tested on the Stanford-Binet upon entry, and again 4 years-9 months later. Their mean growth in standard score units (i.e., IQ) was assessed for statistical and educational significance (Tallmadge, 1977) after a correction for regression was made (Campbell & Erlebacher, 1970). The mean IQ at entry was 41.9 (SD = 2.6). After five years of the program, the mean IQ was 50.6 (SD = 5.4). If there was only a regression effect operating, the mean IQ at post-test would have been 44.8; there was a gain of .36 standard deviation units beyond what would be expected from regression. This gain was therefore found both statistically significant and educationally significant. The moderately retarded children in this sample were gaining at a significantly faster rate than their non-handicapped peers, even though typically these

children would be expected to lose against the norm group. These students were also tested on Australian standardized tests in Reading and Language; the mean scores were comparable to an end-of-third grade level in language and early third grade level in reading. The children appeared to be close to mastering the basic functional literacy skills essential for vocational training and independent living.

Ideally, the progress of these students would have been compared to a norm group of moderately handicapped students; but such data is currently not available. The standardization sample of a well-known test such as the Stanford-Binet or the Comprehensive Test of Basic Skills is a legitimate, if overly stringent, comparison standard, and one that is often utilized in special education evaluation. It makes special sense for samples of mildly handicapped students where "normalization" in academic areas is an immediate goal.

Time-Series Design: Effects of Direct Instruction Programs on Mainstreamed Low IQ Students. The following norm-referenced comparison is also an example of a longitudinal, time-series design. This design uses repeated observations, or measurements of the same class of behavior (in this case, reading and math performance) at specific points in time. Gersten, Becker, Heiry, and White (Note 5) followed the annual progress of over 900 low income students in the Direct Instruction Follow Through program for four full years, from kindergarten through third grade. These students came from twelve communities--inner city areas such as New York and Washington, D.C., the Rosebud Sioux reservation in South Dakota, and rural communities such as Flippin, Arkansas. The purpose of the analyses was to look at the effect of Direct Instruction on students entering with low cognitive skills (i.e., low IQs). Gersten et al. (Note 5) looked at the yearly achievement test profiles for low income Follow Through

children who entered with IQs on the Slosson Intelligence Test of 70 or below, and compared them to students entering with IQs in the normal range. The low IQ children would normally qualify for special education services, but because of the model's belief that students should be in regular classrooms, the children were de facto "mainstreamed." Inservice training for the teachers and paraprofessionals involved in the programs emphasized that teachers were responsible for the growth of all children, including the lower performers, and that with adequate practice, feedback, and reinforcement, the low performers could learn at a normal rate. IQs were never used for placement purposes or referral for special services; placement and grouping were determined solely by criterion-referenced tests. The Metropolitan Achievement Test and Wide Range Achievement Test were used to assess progress in reading and mathematics. The Slosson was used to measure growth in general language competence.

In analyzing the longitudinal achievement data, the hypothesis was that the yearly academic growth rate would be no different for the lower performers than the others in the program. Though the low IQ students enter kindergarten with lower academic skill levels, and leave the third grade with lower skill levels than their normal or high IQ peers, we believed that their yearly growth, their ability to profit from schooling, would be roughly the same as their peers.

No significant interactions were found between entry IQ level and WRAT Reading (a measure of word identification) or Total Math on the Metropolitan Achievement Test (a measure of math problem solving, concepts, and computation). Considering the large sample sizes ($N=594$) and consequent high statistical power, there is reasonably clear evidence of a

Insert Figures 1, 2 and 3 about here

non-significant effect. Figures 1 and 2 depict yearly gains for the low IQ block (70 and below), those with IQs between 71 and 90, and the higher IQ blocks on WRAT Reading and MAT Math. On WRAT Reading, the low IQ students are virtually at norm levels (the 47th percentile) by the end of kindergarten and continue to make slow but steady growth. By the end of third grade, performance corresponds to the 70th percentile, or a 4.3 grade level. A similar pattern is found for the larger 71-90 IQ group. For MAT Math, the growth rate for both the below 71 and 71-90 IQ blocks corresponds to 1.0 Grade Equivalent unit for each year in school (see Figure 2).

The only significant interaction was found on the MAT Total Reading test (Elementary level) during the third grade. MAT Total Reading is primarily a measure of reading comprehension and vocabulary; here, the low IQ group sample gained significantly less than the higher IQ children during third grade. (See Figure 3.)

The average gain on the Slosson Intelligence Test, a shortened version of the Stanford-Binet, was 8 IQ points, maintained over the entire four years, for the entire sample of 717 children. The mean growth for the lower IQ block is 17 IQ points over the 4-year period.

These low IQ students are demonstrating one year's growth for each year in school in mathematics, and even higher levels in reading (decoding). They also show growth in the general language and conceptual skills assessed by the Slosson Intelligence Test, representing a net gain of 17 IQ points. The picture is less promising in the area of reading comprehension and Vocabulary. Although they seem to adequately learn the initial decoding skills, these children need more systematic instruction in vocabulary concepts and word meanings as they progress through school (Becker, 1977). However, in general, these secondary analyses document the effectiveness

of a mainstreamed model with Direct Instruction teaching techniques and materials and mastery learning principles for children who would be typically considered "at risk" or mildly handicapped.

Documentation Activities

At this point, I wish to distinguish between evaluation research, where the effectiveness of an instructional program or instructional variable is assessed using a valid research design, and documentation activities, where test scores describing the impact of a program are recorded without the benefit of the comparison standards present in experimental or quasi-experimental designs. In the last decade there has been a good deal of documentation of the impact of Direct Instruction programs with handicapped children, in Australia. These studies are summarized in ~~unmarized~~ in Becker, Engelmann, Carnine, and Maggs (1981), Maggs and Maggs (1979), and Stevens (1980). Though this activity may offer administrators, decision makers, curriculum programmers, and parents some evidence of the impact of these programs, rarely can serious inferences be drawn. An example will illustrate this situation.

Appfel, Kelleher, Lilly and Richardson (1975) documented the effectiveness of two structured reading programs--Distar and Rebus (Woodstock, 1967)--for 60 moderately retarded children. They wished to supply evidence that given appropriate curricula material, many moderately retarded children can be taught to read on a functional level. The Rebus and Distar reading programs were selected because both were structured programs. However, Rebus utilized a whole word approach, whereas Distar utilized a synthetic phonics approach to oral group instruction. The authors did not use random assignment, nor did they indicate on what basis subjects were

assigned to a reading program. No pretest information was provided. Thus, unlike the L'E. Stein and Goldman study, the reader has no idea whether the two samples were comparable before instruction began.

Student progress was assessed by individually administered criterion-referenced measures. While this procedure is fine for formative evaluation, it does not allow one to compare the relative effectiveness of the two programs. Nor, on the basis of this study, can one assess whether one type of program worked better for a particular type of student. Yet the project did document that nearly all the young THR participants demonstrated some capacity to profit from reading instruction. A few failed, but some students in both Distar and Rebus advanced to more conventional reading programs. Using Lindsley's (1964) graphic technique for analysis of progress rate, the authors concluded that 20 of the 31 Distar students made Good Progress (65%) and the remaining 11 poor progress. For the Rebus sample, 6 of the 23 (26%) were considered to be making Good Progress and 11 of the 23 (48%) to be making poor progress. (The remaining were considered "neutral.") Despite some confusing methodology, this study uses teacher-collected progress data to document that reading can be taught to many of the moderately retarded participants with a reasonably high success rate. It shows that Direct Instruction procedures are effective 65 percent of the time, and the success rate for the whole-word approach is appreciably lower, as one might expect on the basis of research with non-handicapped learners (Pflaum, Walberg, Karregianes, & Rahser, 1981).

Discussion

A reasonably large number of studies have shown that Direct Instruction reading and language programs consistently produce higher academic gains

than traditional approaches in both mainstreamed settings (Gersten et al., Note 5) and self-contained classrooms (e.g., Lloyd et al., 1981; Maggs & Morath, 1976) across a range of handicapping conditions. The L' E. Stein and Goldman study indicated that some of the more subtle principles of Direct Instruction theory, such as the insistence of complete (rather than partial) mastery in each stage of the learning process, may be quite important for special education students. There is a clear need for further comparative research and norm-referenced evaluations of these programs across a greater range of handicapping conditions and academic areas.

Several conceptual and related methodological issues need to be considered in future research. Gowin and Millman (1981) recently stated, "evaluative inquiry should raise questions, and produce clarifications (not answers)..." (p. 85). The studies discussed in this paper raise many subtle, important questions and issues. First and foremost is the need for increasingly precise definition and measurement of the independent variable--Direct Instruction.

It is unclear from the studies reviewed in this paper whether the increases in academic performance were due to (a) the curriculum programs utilized, (b) the teacher behaviors called for in the programs, (c) the or a combination of these factors. None of the studies actually assessed whether the programs were actually implemented as specified. The authors are rather vague in describing how the programs were adapted to meet the needs of the moderately and severely handicapped learners. Furthermore, with the exception of the L'E. Stein and Goldman (1980) study, relatively little was specified (let alone measured) in the comparison classrooms.

A newer line of research, called instructional dimensions research (Leinhardt, Zigmond, & Cooley, 1981) or explanatory observational studies (Cooley, 1978) can also lend important information about exactly which teaching practices or aspects of the curriculum are most effective. For this type of study, observers look at a set of variables of interest-- e.g., degree of structure, allocation of academic time, student engagement during lessons, student accuracy rate during lessons--in a large number of classrooms utilizing a range of instructional approaches. The researcher attempts to detect which variables, or combination of variables, are linked to academic gains. There is no need to select a comparison group; the design capitalizes on the natural variation which is everpresent in education. Therefore, these studies need not be bound by traditional achievement tests. As long as the same test is used in all classrooms, criterion-referenced tests tailored to the instructional programs actually used in the classroom can be used.

A recent study by Leinhardt, Zigmond, and Cooley (1981) with 105 learning disabled students exemplifies this approach. The authors found that three teacher behaviors were significantly related to gains in reading-- use of reinforcers, cognitive press (the degree to which the teacher focused the child on academic material), and teacher instruction (presentation of models, explanation, provision of feedback). All three are critical in most definitions of Direct Instruction (e.g., Reith et al., 1982). This study was the first to demonstrate the importance of these behaviors in teaching handicapped students. Further explanatory observational studies can be conducted that look at more specific teacher behaviors, such as correction procedure, type of teacher questions, presentation rate, and

the nature of the instructional formats. Such studies can contribute to identifying the critical features of Direct Instruction.

Need for Formative Research. Another area for future programmatic research is continuing formative evaluation studies, such as those conducted by Carnine, 1981; Close, Taylor, Larrabee, & Taylor, Note 6). Gowin and Hillman (1981) state that future evaluation research "designs are unlikely to be massive and confirmatory, but rather small and exploratory" (p. 85). In these small scale studies, only one instructional variable is manipulated.

Several studies with handicapped subjects have isolated a single component of Direct Instruction as the independent variable. To date, research has shown the superiority of teaching sequences that separate the introduction of visually or auditorily similar material (such as "b-d" or "p-q") over those that attempt to introduce both members of the set at once (Carnine, 1981). White (Note 7) documented how training sequences designed for normal learners must be simplified and adapted for handicapped learners, who are confused by too much variation in irrelevant stimulus features when learning a new geometric concept. Gersten et al. (in press) demonstrated that, when teaching language concepts to handicapped students, dynamic presentations using manipulations of real objects are markedly superior to teaching from a series of line drawings. Close et al. (Note 6) demonstrated that when teaching mildly retarded adults how to cash a check, learning was more efficient if the instructor used a correction procedure that provided the learner with an explicit strategy and immediate practice in application of the strategy.

Single subject designs can also be extremely useful in formative evaluation, especially when dealing with moderately and severely handicapped

individuals. Colvin and Horner (in press) used a multiple baseline design to investigate the effect of the range of examples used in vocational training of severely retarded adolescents on their ability to generalize the skills.

Summary. Generally, a quite optimistic picture has emerged from these studies. In a recent review of research on Direct Instruction, Cotton & Savard (1982) concluded:

Direct instruction, as that term is used to denote the agreed-upon set of teaching strategies and behaviors described earlier, is very effective for promoting basic skill development among students generally. When teachers set and articulate learning goals, offer highly structured lessons, ask questions which are specific and narrow in scope, provide corrective feedback, and communicate affection and support to students, achievement results are superior to those obtained with other, less direct methods...

Distar programs are also very effective for instructing low-ability children and educable special education students in the primary and upper elementary grades.

The next decade can provide increasingly clear specification of which components of Direct Instruction are most critical for success with handicapped individuals, and precisely how programming principles and teaching strategies should be adapted to meet the needs of this population.

Reference Notes

1. Kennedy, H. (Chair). Minutes of the proceedings of the Special Interest Group on Special Education Research. American Education Research Association; Boston, Mass., 1980.
2. Decker, M. C., Gersten, R., & Carnine, D. Design and measurement issues in Follow Through evaluation research. Final report submitted to Office of Education, DHEW, pursuant to contract RFP 78-101, 1979.
3. Lloyd, J. Personal communication, 1981.
4. Maggs, A. Personal communication, 1980.
5. Gersten, R., Decker, M. C., Heiry, T. J., & White, W. A. T. Relationship between entry IQ and yearly academic gains in a direct instruction model. Paper presented at AERA; Boston, Mass., April 1981.
6. Close, D., Taylor, V., Larrabee, D., & Carlson, C. The effects of varied correction procedures on the acquisition and retention of complex information by mildly retarded adults. Unpublished manuscript. University of Oregon, 1982.
7. White, W.A.T. The effect of decreasing auditory similarity in initial teaching of figure/label correspondences to handicapped learners. Unpublished masters thesis. University of Oregon, 1979.

References

- Appfel, J. A., Kelleher, J., Lilly, M. S., & Richardson, R. Developmental reading for moderately retarded children. Education and Training of the Mentally Retarded, 1975, 10, 229-235.
- Bateman, B. D. Teaching reading to learning disabled children: A fourth approach. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.
- Bateman, B., & Carnine, D. W. Direct instruction DISTAR. In H. G. Harring & B. Bateman, Teaching the learning disabled child. Englewood Cliffs, NJ: Prentice Hall, 1977.
- Becker, W. C. Teaching reading and language to the disadvantaged--What we have learned from field research. Harvard Educational Review, 1977, 47, 518-545.
- Becker, W. C., & Engelmann, S. Teaching 3: Evaluation of instruction. Chicago: Science Research Associates (SRA), 1976.
- Becker, W. C., Engelmann, S., Carnine, D., & Maggs, A. Direct instruction technology--Making learning happen. In P. Karoly & J. J. Steffen (Eds.), Advances in child behavior analysis and therapy (Vol 2). New York, Gardner Press, 1981.
- Becker, W. C., Engelmann, S., Carnine, D., & Rhine, R. The direct instruction model. In Rhine, R. (Ed.), Encouraging change in America's schools. A decade of experimentation. New York: Academic Press, 1981.
- Bellamy, G. T., Horner, R. H., & Inman, D. P. Vocational rehabilitation of severely retarded adults: A direct service technology. Baltimore: University Park Press, 1979.
- Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multifruit-multi-method matrix. Psychological Bulletin, 1959, 56, 81-105.
- Carnine, D. W. Reducing training problems associated with visually and auditorily similar correspondences. Journal of Learning Disabilities, 1981, 14, 276-279.
- Carnine, D., & Gersten, R. Effective mathematics instruction for low income students: Results of longitudinal field research in twelve districts. Journal of Research on Mathematics Education, 1981, 13, 145-152.

- Chall, J. S. The great debate: Ten years later, with a modest proposal for reading stages. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.
- Charters, W. W., & Jones, J. On the risk of appraising non-events in program evaluation. Educational Researcher, 1973, 2(11), 5-7.
- Colvin, G. T., & Horner, R. H. Experimental analysis of generalization: An evaluation of a general case program for teaching motor skills to severely handicapped learners. In J. Hogg & P. J. Miffler (Eds.), Advances in mental handicap research. Vol 2. Sussex, UK: John Wiley & Sons, in press.
- Cook, T. D., & Campbell, D. T. Quasi-experimentation: Design and analysis for field settings. Chicago: Rand McNally College Publishing Company, 1979.
- Cooley, W. W. Explanatory observational studies. Educational Researcher, 1978, 7(9), 9-15.
- Cotton, K., & Savard, W. Direct instruction: Research on school effectiveness project. Portland, OR: Northwest Regional Laboratory, 1982.
- Engelmann, S., Becker, W. C., Hanner, S., & Johnson, G. Corrective reading: Series guide. Chicago: Science Research Associates, 1978.
- Gall, M. D. The importance of context variables in research on teaching skills. Research in Teacher Education, 1977, 28, 43-48.
- Gersten, R., Carnine, D., & Williams, P. Measuring implementation of a structured educational model in an urban setting: An observational approach. Educational Evaluation and Policy Analysis, 1982, 4, 67-79.
- Gersten, R., & Maggs, A. Teaching the general case to moderately retarded children: Evaluation of a five-year project. Analysis and Intervention in Developmental Disabilities, in press.
- Gersten, R., White, W. A., Falco, R., & Carnine, D. Teaching basic discriminations to handicapped and non-handicapped individuals through a dynamic presentation of instructional stimuli. Analysis and Intervention in Developmental Disabilities, in press.
- Gibson, E. J., & Levin, H. The psychology of reading. Cambridge, MA: Massachusetts Institute of Technology Press, 1975.
- Gill, E. J. Methods of teaching reading: A comparison of results. In C. Walcutt, J. Lampart, & T. McCracker (Eds.), Teaching reading: A phonic/linguistic approach to developmental reading. New York: MacMillan, 1974.

- Gilmore, J. V., & Gilmore, E. C. Gilmore oral reading test. New York: Harcourt Brace Jovanovich, 1968.
- Good, T., & Grouws, D. The Missouri Mathematics Effectiveness Project. Journal of Educational Psychology, 1979, 71, 355-362.
- Gowin, D. B., & Millman, J. Review of Toward reform of program evaluation. Educational Evaluation and Policy and Analysis, 1981, 3, 85-87.
- Hall, G. E., & Loucks, S. A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 1977, 14, 264-276.
- Horst, D. P., Tallmadge, G. K., & Wood, C. T. A practical guide to measuring project impact on student achievement. Monograph No. 1 on Evaluation in Education. Washington, D.C.: U.S. Government Printing Office, 1975.
- House, E. R., Glass, G. V., McLean, L. D., & Walker, D. E. No simple answer: Critique of the "Follow Through" evaluation. Harvard Educational Review, 1978, 48, 128-160.
- Irvin, L., Gersten, R., Bellamy, G. T., Taylor, V., & Close, D. The trainee performance sample: Improving vocational skill assessment for the severely retarded. American Journal of Mental Deficiency, 1981, 85, 631-633.
- Jastak, J. F., & Jastak, S. R. The Wide Range Achievement Test. Wilmington, Del.: Guidance Associates of Delaware, Inc., 1965.
- Kephart, H. C. The perceptual motor aspect of learning disabilities. Journal of Exceptional Children, 1964, 31, 21-206.
- L'E. Stein, C., & Goldman, J. Beginning reading instruction for children with minimal brain dysfunction. Journal of Learning Disabilities, 1980, 13, 219-222.
- Lindsley, O. P. Direct measurement and prosthesis of retarded behavior. Journal of Education, 1964, 147, 62-81.
- Lloyd, J., Cullinan, D., Heins, E. D. Epstein, M. H. Direct instruction: Effects on oral and written language comprehension. Learning Disabilities Quarterly, 1980, 3, 70-76.
- Lloyd, J., Epstein, M. H., & Cullinan, D. Direct teaching for learning disabilities. In J. Gottlieb & S. Strichart (Eds.), Perspectives on handicapping conditions: Current research and application in learning disabilities. Baltimore, MD: University Park Press, 1981.
- Maggs, A., & Maggs, R. K. Direct instruction in Australia. Journal of Special Education Technology, 1979, 3, 26-34.

- Maggs, A., & Morath, P. Effects of direct verbal instruction on intellectual development of institutionalized moderately retarded children: A 2-year study. Journal of Special Education, 1976, 10, 357-64.
- Matarazzo, J. D. Wechsler's measurement and appraisal of adult intelligence. New York: Oxford University Press, 1972.
- Meyer, L. A. An experimental study of the treatment effects of word-analysis and word-supply correction procedures during word attack. Reading Research Quarterly, in press.
- Oregon Mental Health Division. Student Progress Record. Salem, OR: Author, 1977.
- Pflaum, S. W., Wahlberg, H. J., Karegianes, M. L., & Rahser, S. P. Reading instruction: A quantitative analysis. Educational Researcher, 1980, 9, 12-18.
- Reith, H. J., Polsgrove, L., & Semmel, M. I. Instructional variables that make a difference: Attention to task and beyond. Exceptional Education Quarterly, 1982, 2, 61-71.
- Rosenshine, B. V., & Berliner, D. C. Academic engaged time. British Journal of Teacher Education, 1978, 4, 3-16.
- Serwer, B. L., Shapiro, B. J., & Shapiro, P. S. The comparative effectiveness of four methods of instruction on the achievement of children with specific learning disabilities. Journal of Special Education, 1973, 7, 241-249.
- Slosson, R. L. Slosson intelligence test. East Aurora, NY: Slosson Educational Publications, 1963.
- Stebbins, L., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. Education as experimentation: A planned variation model, Vols. IVA-D. Cambridge, Mass., 1977.
- Stephens, E. Direct instruction research with moderately and severely handicapped in Australia: A review. ASET Journal, 1980, 12, 41-46.
- Stevens, R., & Rosenshine, B. Advances in research on teaching. Exceptional Education Quarterly, 1981, 2, 1-10.
- Tallmadge, G. K. JDRP ideabook. Washington, D.C.: U.S. Government Printing Office, 1977.
- White, O. R. Adaptive performance objectives: Form versus function. In W. Sailor, B. Wilcox, & L. Brown (Eds.), Methods of instruction for severely handicapped students. Baltimore: Paul H. Brookes, 1980.
- Woodcock, R. W. Peabody rebus reading program. Circle Pines, MI: American Guidance Service, 1967.

Figure Captions

Figure 1. WRAT Reading: Longitudinal progress by IQ block for low-income children in Direct Instruction Follow Through program.

Figure 2. MAT Total Mathematics: Longitudinal progress by IQ block for low-income children in Direct Instruction Follow Through.

Figure 3. MAT Total Reading: Longitudinal progress by IQ block for low-income children in Direct Instruction Follow Through program.

